Improving Polynucleotide Ligation Reactions

Field of Invention

The present invention relates to methods for improving polynucleotide ligation reactions.

Background to Invention

Berg and Boyer created the first recombinant DNA molecule in 1972. This simple concept of recombination - the splicing together of two pieces of DNA and fusing them by ligation, is the basis for the entire field of molecular biology. Molecular biology has become ubiquitous to the point where it is central to the majority of all biological research. The ligation reaction is performed thousands of times a day in research and diagnostic laboratories worldwide. Given the boundless opportunity presented by genetic engineering, the ligation reaction is likely to remain a central technique for many years to come.

The ligation reaction itself is chemically simple, comprising the linking of two nucleotides by the creation of a phosodiester bond between the 3' hydroxyl of one nucleotide and the 5' phosphate of another, by a ligase enzyme. There are two types of ligation, known as "sticky end" and "blunt end", depending on the presence or lack (respectively) of complementary single stranded regions on the two polynucleotides to be joined, in proximity to the ligation location. "Sticky-end" ligations involve the hybridisation of complementary single stranded sequences between the two polynucleotides to be joined, prior to the ligation event itself. Sequences that have similar but not 100% complementary single stranded sequences will still be ligated, known as a mismatch ligation. These result in the ligation of an incorrect sequence and decrease the efficiency and fidelity of the overall ligation reaction.

Since ligation is such an important reaction, ligases are available on the market that are improved, modified and optimised to give maximum efficiency. These enzymes are expensive and it is therefore desirable to use as a small amount as is possible without reducing the efficiency of the reaction and whilst avoiding mismatch ligation. Mismatch ligations are problematic as they are

deleterious to the fidelity of the ligation process. It is therefore desirable to minimise mismatch ligations.

Current methods of increasing ligation specificity include decreasing the amount of ligase and increasing the salt in the reaction mix to slow down the reaction. Since match ligations are much faster than mismatch ligations, the increased specificity observed using this technique is a result of the slower reaction speed and whilst this increases the match: mismatch ratio, it results in a low yield and does not prevent mismatch ligations.

There is therefore a need for improvements in ligation reactions.

Summary of Invention

The present invention is based on the realisation that specificity in "sticky-end" ligations can be increased by including short adapters that reduce the occurrence of mismatch ligation.

According to a first aspect of the invention, a method for improving the specificity of a ligation reaction carried out between a first double stranded polynucleotide having a single stranded portion and a second polynucleotide having a complementary single stranded portion, said second polynucleotide being present in a sample comprising a mixture of different polynucleotides, comprises:

contacting the sample, under hybridising conditions, with the first polynucleotide and one or more third polynucleotide(s), wherein the third polynucleotide(s) comprises a single stranded portion that differs from the single stranded portion of the first polynucleotide by at least one base substitution, and carrying out a ligation reaction.

The present invention improves the yield of match ligations by reducing mismatch ligations through the use of blocking polynucleotides which hybridise to incorrect single stranded overhangs on the second polynucleotides.

Description of the Drawings

The present invention is illustrated by reference to the accompanying drawing, where:

Figure 1 is a graphic illustration of match:mismatch ratio as a function of time, wherein Figure 1a illustrates the ratio in the absence of blocking adapters, and Figure 1b illustrates the ratio in the presence of blocking adapters.

Detailed Description of the Invention

The present invention is used to increase specificity of polynucleotide ligations. The term "polynucleotide" is used herein to refer to biological molecules made up of a plurality of nucleotides. Preferred polynucleotides include DNA, RNA and synthetic analogues thereof, including PNA.

The term "hybridising conditions" is used herein to refer to conditions that allow complementary base pairing to occur between two polynucleotides, such that two complementary single stranded polynucleotides will hybridise to form a duplex. Such conditions are well known in the art. An Example of such conditions is incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C. Alternative conditions will be apparent to the skilled person and are applicable to the present invention.

In any ligation reaction, two polynucleotide molecules are joined. The term "first polynucleotide" is used herein to refer to one of the two intended targets of ligation. The term "second polynucleotide" is used herein to refer to the other of the two intended targets of ligation.

A non-limiting example of the terms "first polynucleotide" and "second polynucleotide" comprises the "first polynucleotide" being a DNA vector into which an insert, the "second polynucleotide", is to be ligated to form a recombinant construct.

In any ligation reaction, there may be polynucleotides present which are neither "first polynucleotides" or "second polynucleotides", in the sense that they are not intended to be part of the ligation reaction. These polynucleotides interfere with the ligation between the first and second polynucleotides, which results in mismatch ligations.

As used herein, the term "third polynucleotide" is used to describe polynucleotides which are added to the ligation reaction mixture to hybridise to any polynucleotide which is not a first or second polynucleotide, preventing the unwanted polynucleotides from reacting with the other components of the reaction mix. The third polynucleotides are not totally complementary to the first or second polynucleotides.

The method increases specificity in polynucleotide ligations through the addition of one or more third polynucleotide(s) into a reaction mix. This reaction mix comprises a first polynucleotide and a second polynucleotide, which contain complementary single stranded portions. The second polynucleotide is present in a sample comprising a mixture of different polynucleotides, and is the intended target for binding to the first polynucleotide. The third polynucleotide(s) comprises at least a single stranded portion that differs from the single stranded portion of the first polynucleotide by at least one base. The number of differences between the first and third polynucleotides may depend on the size of the single stranded portions involved. For example, if the single stranded portion is only 3 bases in length, a single difference may be suitable, but if the single stranded portion is 6 bases in length, multiple differences may be preferred. The differences may be substitution(s), deletion(s) or addition(s).

The third polynucleotide may be added to the sample containing the second polynucleotide simultaneously with or sequentially before or after the first polynucleotide. The third polynucleotide is preferably added to the sample containing the second polynucleotide, along with the first polynucleotide.

Preferably, the third polynucleotide is present in excess with respect to the first and second polynucleotides, to ensure that all other polynucleotides in the sample are hybridised by the third polynucleotide.

It is intended that the first and second polynucleotides hybridise and are ligated together, to the exclusion of other polynucleotides in the sample. The third polynucleotide(s) hybridise to the other polynucleotides in the sample which would otherwise compete for binding to the first polynucleotides, effectively preventing them from hybridising to the first polynucleotides and increasing the

number of correct binding events between the first and second polynucleotides. This increases the specificity of the overall ligation reaction.

Preferably the mixture of third polynucleotides comprises double stranded polynucleotides with a single stranded portion, such that the single stranded portion hybridises its complementary region on incorrect first and second target polymers.

Preferably, the single stranded portion of each of the first, second and third polynucleotides is from 3 to 6 bases in length. Most preferably, the single stranded portion is 4 bases in length.

Figure 1 is a graphical representation of the match:mismatch ratio as a function of time. This ratio becomes lower as the reaction progresses, since the match reaction rapidly reaches plateau and is caught up by the slower mismatch reaction. Traditional methods of increasing specificity merely slow the reaction (using less ligase or increasing salt concentration) and shift the reaction to the left of the graph, where the match:mismatch ratio is favourable but yield is decreased. The present invention ensures that the mismatch ligations do not increase. The match reaction can proceed to full term without the mismatch reaction ever catching up. This provides an optimised match:mismatch ratio.

The invention will now be illustrated with reference to the following, non-limiting, example.

<u>Example</u>

A test system was set up to measure the effect on ligation specificity of adding blocking adapters to a ligation reaction. The goal of this test system was to measure ligation specificity, i.e. the percentage of correctly ligated molecules relative to the total number of molecules ligated. The test system used makes use of the fact that a ligation product containing a single base mismatch in the ligation overhang region differs from correctly ligated ligation product by only one base. The protocol used was developed to measure ligation specificity (match ligation as percentage of total ligation) using the Homogeneous Mass Extend method (Sequenom) and the MassARRAY stystem (Sequenom). The protocol was used to measure the specificity of ligation for the 3'-most base (upper strand) of the ligation overhang region.

The polynucleotides used in this Example are shown in Table 1.

Table I: Sequences of adapters used

DPA 12

5′-
TGTGTCCGCGTGGCTCTTCTATTCTTGGCTTTTCGTCGCTTGGCTTTTCGTCGCTTGGTCATTCGTCGCTTGG
TCATTCGTCGCTTGGCTTTTCGTCGCTTGGTCATTCGTCG -3′ (SEQ ID NO: 1)
3′-
ACACAGGCGCACCGAGAAGATAAGAACCGAAAAGCAGCGAACCGAAAAGCAGCGAACCAGTAAGCAGCGAACC
AGTAAGCAGCGAACCGAAAAGCAGCGAACCAGTAAGCAGCTAA -5′ (SEQ ID NO: 2)

DPA 14

5′-
TGTGTCCGCGTGGCTCTTCTATTCTTGGCTTTTCGTCGCTTGGCTTTTCGTCGCTTGGTCATTCGTCGCTTGG
TCATTCGTCGCTTGGTCATTCGTCGCTTGGCTTTTCGTCG -3′ (SEQ ID NO: 3)
3′-
ACACAGGCGCACCGAGAAGATAAGAACCGAAAAGCAGCGAACCGAAAAGCAGCGAACCAGTAAGCAGCGAACC
AGTAAGCAGCGAACCAGTAAGCAGCGAACCGAAAAGCAGCTAA -5′ (SEQ ID NO: 4)

DPA 24

5′-
TGTGTCCGCGTGGCTCTTCTATTCTTGGCTTTTCGTCGCTTGGTCATTCGTCGCTTGGTCATTCGTCGCTTGG
CTTTTCGTCGCTTGGCTTTTCGTCGCTTGGCTTTTCGTCG -3′ (SEQ ID NO: 5)
3′-
ACACAGGCGCACCGAGAAGATAAGAACCGAAAAGCAGCGAACCAGTAAGCAGCGAACCAGTAAGCAGCGAACC
GAAAAGCAGCGAACCGAAAAGCAGCGAACCGAAAAGCAGCTAA -5′ (SEQ ID NO: 6)

DPA 44

5′-
GTGTCCGCGTGGCTCTTCTATTCTTGGTCATTCGTCGCTTGGCTTTTCGTCGCTTGGTCATTCGTCGCTTGGT
CATTCGTCGCTTGGCTTTTCGTCGCTTGGTCATTCGTCG -3′ (SEQ ID NO: 7)
3′-
ACACAGGCGCACCGAGAAGATAAGAACCAGTAAGCAGCGAACCGAAAAGCAGCGAACCAGTAAGCAGCGAACC
AGTAAGCAGCGAACCGAAAAGCAGCGAACCAGTAAGCAGCTAA -5′ (SEQ ID NO: 8)

Target

5′- ATTTATCTGCTGCATGATCCGATAGTGCGAAT -3′
3′- ATAGACGACGTACTAGGCTATCACGCTTANNNN -5′ (SEQ ID NO: 9)

SLA 13

5′- NATCTAGATGCACTCCCGGACCTC -3′ BLA 13-1
3′- ATCTACGTGAGGGCCTGGAG -5′
(SEQ ID NO: 10)

SLA 14

5′- NATGTAGATGCACTCCCGGACCTC -3′
3′- ATCTACGTGAGGGCCTGGAG -5′
(SEQ ID NO: 11)

SLA 24

5′- NCGATAGATGCACTCCCGGACCTC -3′
3′- ATCTACGTGAGGGCCTGGAG -5′
(SEQ ID NO: 12)

SLA 44

5′- NGTATAGATGCACTCCCGGACCTC -3′
3′- ATCTACGTGAGGGCCTGGAG -5′
(SEQ ID NO: 13)

5′- NBTCATGAGCTGGGCGGCACGTAT -3′
3′- TACTCGACCCGCCGTGCATA -5′
(SEQ ID NO: 14)

BLA 13-2

5′- NAVCATGAGCTGGGCGGCACGTAT -3′
3′- TACTCGACCCGCCGTGCATA -5′
(SEQ ID NO: 15)

BLA 13-3

5′- NATDATGAGCTGGGCGGCACGTAT -3′
3′- TACTCGACCCGCCGTGCATA -5′
(SEQ ID NO: 16)

In Table I, "N" represents any of the bases G, C, T and A; "B" represents any of the bases G, C and T; "D" represents any of the bases A, C and T; "H" represents any of the bases A, C and T; and "V" represents any of the bases A, G and C.

Reactions were carried out where three of the following adapters were ligated together (see Table I):

1) a Design Polymer Adapter (DPA)

2) a target molecule with a 4 nt 5' overhang representing all 256 possible permutations of 4 nucleotides ('NNNN')

3) a Specific Ligation Adapter (SLA) meant to ligate specifically to a subset (1/64th) of the target molecules.

Based on its 4 nt overhang, the SLA was supposed to ligate to only 4 out of the 256 permutations of the target.

There were four variants of each DPA (designated 13, 14, 24 and 44) and four of each SLA (designated similarly), see Table I. In some cases, a specific set of three Blocking Adapters (BLA's) were added to each ligation reaction to block all nine single base mismatches possible for a particular reaction (see Table I); there were 3 x 4 variants of each BLA (Table I).

For each reaction, DPA, Target and the corresponding SLA were ligated (at a molecular ratio of DPA:Target:SLA = 64:64:10) in the presence or absence of the specific set of nine BLA's (each BLA was at an equimolar ratio relative to the SLA) using T4 DNA ligase (Fermentas). PCR using the primers described in Table II was performed on 1/20th of the ligation reaction to amplify the ligation products.

Table II: PCR and Extension primers (in 5' to 3' orientation)

Forward PCR primer	ACGTTGGATGTGTGTCCGCGTGGCTCTTCT (SEQ ID NO: 17)
Reverse PCR primer	ACGTTGGATGATGGGCTTTTGAGGTCCGGGAGTG (SEQ ID NO 18)
Extension primer	GAGGTCCGGGAGTGCATCTA (SEQ ID NO: 19)

Targets ligated to BLAs could not be amplified during this PCR due to the different sequence at their 3' end (upper strand). PCR products were isolated and concentrated using the MinElute PCR Cleanup Kit (Qiagen).

The cleaned, concentrated PCR products were used in an extension reaction following the Homogenous Mass Extend protocol described in the Mass ARRAY User's Manuals (Sequenom). Basically, the reaction consisted of the product of the PCR amplified ligation product, an extension primer complementary to the sequence 5' of the base to be investigated (see Table II), a 'Stop mix' (a specific mixture of equimolar amounts of one NTP in the dNTP form and the remaining three NTPs in the ddNTPs form), a thermostable

polymerase in 1 x buffer (all components from Sequenom except: ddNTPs from Roche, dNTPs from Amersham). The stop mix was chosen so that a base at the most 3' end of the ligation overhang region resulting from a correct (match) ligation would yield a 2-base extension product whereas a base resulting from a mismatch ligation would yield a 1-base extension product.

Extension reactions and the subsequent washes were performed following the Homogenous Mass Extend method (Sequenom). Extension products were spotted on a SpectroCHIP (Sequenom) and separated according to mass as described in the Mass ARRAY User's Manuals (Sequenom). The results were a set of peaks representing unextended extension primer and (1-base and 2-base) extension products. The intensities of these peaks were transferred manually to a spreadsheet. The relative intensities of the 2-base extension product (representing the match ligation extension product) versus the total intensity of the extension products was calculated.

The results of the experiments are shown in Table III.

Table III: Percentage match ligation extension product of the total extension product for the 3'-most base of the overhang

Reaction (overhang)	Without Blocking Adapters	With Blocking Adapters
13 (NATC)	84.4 %	88.0 %
14 (NATG)	100.0 %	100.0 %
24 (NCGA)	100.0 %	100.0 %
44 (NGTA)	75.8 %	84.4 %

These results indicate that two out of the four ligation reactions (numbers 14 and 24) did not produce a detectable peak for a mismatch ligation product at the 3'-most position. For the other two reactions (13 and 44) the peak intensities for the match-ligation extension product detected represented 84.4% (reaction 13) and 75.8% (reaction 44) of the total extension product intensity, respectively, in the reactions without Blocking Adapters. These intensities were higher when

Blocking Adapters were included in the reactions: 88.0% (reaction 13) and 84.4% (reaction 44). This indicates that blocking adapters helped to reduce mismatch ligations occurring at the 3'-most position.